Capturing Some Lexical Differences between English and Korean: A Generative Lexicon Approach*

Yae-Sheik Lee · Jong-Bok Kim

1. Data

There are several types of language differences between English and Korean, in particular, those differences in conceptual scope differences, lexical gaps, aspectual meaning, and so forth.

1.1. Conceptual Scope Differences

One clear set of facts that differentiate English and Korean is cases where the counterpart words have different denotational scopes. For example, consider the pairs in (1).

(1) < wider scope English words, narrower scope Korean Words >

Nouns	< friend, chinkwu >, < brother, namdongsayng >, < uncle, oysamchon >, < paper, congi >,...
Verbs	< bake, kwupta >, < break, kkayta/pwuswuta >, < wear/put on, ipta >, < tear, ccayta/ccicta >, < cut, ccaluta >, < boil,kkulhita >,...

English nouns and verbs in these examples denote wider conceptual

* An earlier version of this paper, developed from Lee (2000 and b), was presented at the First Seoul International Conference on Discourse and Cognitive Linguistics, June 9-10, 2001. We thank participants in the conference for comments and suggestions. We also are grateful to two anonymous reviewers of this journal for helpful comments and suggestions. We are of course solely responsible for any remaining errors.

categories than those of their Korean counterparts. For example, as represented in (2), the denotations of English *friend* include well-known and likable persons, whereas the Korean *chinkwu* denotes a set of persons who are also of similar age as well as as well-known and likable.

(2) a. friend: well-known, likable
b. Korean chinkwu: well-known, likable, of similar age

The English verb *cut* also has much wider usages than the Korean counterpart *ccaluta*, as shown in (3).

(3) a. cut: to make a narrow opening in (something) with a sharp edge or instrument, accidentally or on purpose:
b. ccaluta: to make a narrow opening on purpose

The Korean word *ccaluta* is used when someone cut something on purpose. This explains the weirdness of (4):

(4) # John-i son-ul ccalassta.
John-Nom hand-ACC cut
'John cut his hand.'

Contrary to the pairs in (2), the pairs in (5) illustrate that English nouns and verbs denote narrower conceptual categories that Korean counterparts.

(5) < narrower scope English words, wider scope Korean words >

Nouns	< behavior, hayngdong >, < medicine, yak >, < glass,can > < chair, uyca >, < bathroom, hwacangsil >, < finger, sonkkalak >, < idea, saynkak >,...
Verbs	< take/ride, thata >, < take, mekta > < crunch, pwuswuta > < hoist, olita >,...

For example, the English word *behavior* denotes the event of doing something in a habitual manner, but the Korean counterpart *hanygdong*

does not include the meaning of 'habitual.'[1]

(6) behavior: doing something habitual
hayngdong: doing something

1.2. Lexical Gaps

There are also lexical gaps between the two languages which may be caused from cultural and social differences.

(7) < No English Word, Korean Word >

Nouns	< ((lit. married woman who looks unmarried), misi >, < (lit. hand-taste), sonmas >, < (lit. sweets, cakes, candies,...), kwaca >,...
Verbs	< (give), tulita >, < (idly walk around), nonilta >, < (not good/miserable), hansimhata >,...

(8) < English word, No Korean >:

Nouns	< default, ?? >, < lap,?? >,...
Verbs	< dust, ?? >, < weed,?? >,...

The data given in (8) show that there are certain verbs that exist only in one of the two languages. Without phrasal paraphrases, it is hard to find one's counterpart in the other language. Such lexical gaps might be due to cultural and social differences imbued in the two languages. Lexical caps can thus be interpreted as conceptual gaps between the speakers of English and Korean or as different degrees in conceptual refinement.

1.3. Lexical Pattern Mismatches

Another disparity we can observe between the two languages concerns denominal verbs. There are no Korean counterparts to English denominal verbs. Instead, Korean employs the combination of a verb and its object,

1. An adjective is required to express this meaning as in *supkwancek hayngdong* 'habitual behavior.'

as illustrated from (9) to (11):

(9) Concept: put X on/in Y:
< butter, pethe-lul paluta >, < asphalt, asuphalthu-lul kkalta >

(10) Concept: put something in X:
< pocket, phokeys-ey netta > < bag, kabang-ey netta >

(11) Concept: get rid of X from somewhere:
< dust, mesci-lul takkanayta >, < weed, capcho-lul ppobanayta >

(12) Concept Remove X from somewhere:
< debone, ppey-lul ceykehata >, < debug, pelley-lul ceykehata >

In these pairs, Korean counterparts denote much more basic concepts than English ones.

The examples of motion verbs also show that English and Korean are different in terms of lexicalizing the concept of 'path' involved in the concept of 'movement.' Korean lexicalizes the path as a verb morpheme; yet English represents it as particle (Talmy 1985, Wienold 1995:306-308):[2]

(13) < No English, Korean >

English	Korean
go/climb up/down G	G-ey/lul olu/nayleyta
go/turn around G	G-lul tolta
put something upon G	G-ey encta

1.4. Aspectual Differences

When English causative verbs are used with past tense, they denote accomplished past situations or events.[3] In contrast, their Korean counterparts

2. There are more examples: English words *bang, beep, clank, crack, hiss, fizzle* are all represented in Korean in the form of *X-hako solina* 'X-with sound.'

3. English causative verbs are such as *attach, burn, chop, deceive, erase, fold, grind, heat, install, kill, lift, mix, open, pull, repair, sharpen, tear, untie, wash, wrap up*, etc.

refer to weak accomplishments in the sense that the accomplishment reading can be cancelled. The contrast between (14) and (15) illustrates this point.

(14) a. #I broke a window, but it wasn't broken.
b. #He started the car, but it didnt' start.

(15) a. Na-nun kewul-ul kkay-ess-ciman kkayecici anh-ass-ta
I-Top mirror-Acc break-Pst-though broken not-Pst-Dec
'I broke the mirror, but it wasn't broken.'
b. Ku-nun sidong-lul kel-ess-ciman kelici-anh-ass-ta.
he-Top start-Acc put on-Pst-but put on-not-Pst-Dec
'He started (his car), but it didn't start.'

We may attribute such a difference to conceptual differences encoded in each relevant lexical entry.

1.5. Polysemous Mismatches

A more interesting set of facts centers on polysemous mismatches. Polysemous lexical items in both languages behave quite differently. Among these, let us consider three main types:

(16) ?? John-un cenyek-ulo yang-ul mek-ess-ta.
John-Top dinner-for lamb-Acc eat-pst-Dec
'John ate lamb for dinner.'

(17) ?? Ku capci/chayk-nun pyencipca-ul haykohay-ss-ta
the magazine/book-Top editor-Acc fire-Pst-Dec
'The magazine/book fired its editor.'

(18) a. ?? Mary-nun chayk-ul sicakhay-ss-ta.
Mary-Top book-Acc begin-Pst-Dec
'Mary began a book.'
b. Mary-nun chayk-ul ilk-kisicakhay-ss-ta.
Mary-Top book-Acc read-ingbegin-pst-Dec.
'Mary began reading a book.'

The examples from (16) to (18) show that even if complementarily polysemous lexical items carry logically related senses, they might not occur cross-linguistically. For example, as shown in (16), English animal-denoting count nouns such as *lamb* can be turned into food-denoting mass nouns. This sense extension process is quite productive in English in a sufficiently marked context as in *John enjoyed the lamb.* (cf. Copestake and Briscoe 1995). But Korean does not allow such a sense extension of animal-denoting nouns, mainly because of the existence of meat-denoting nouns. A similar difference in sense extension can be observed in the concept of the English word *magazine* as given in (17). This word originally denoting a sort of book is extended to denote a company that produces this book. But no such sense extension is possible in Korean. Such lexical discrepancies may result in mismatches in the lexicons of the languages in question.

2. Representing Lexical Concepts

We have seen that the discrepancy data between the two langauges result in the differences in lexical concepts. The question that follows is then how to represent lexical concepts.

Both cognitive linguists and semanticists (e. g. Lakoff 1988, Langacker 1998) accept the view that meaning resides in concept. But one thing to note is that a concept can be only talked about under the assumption that there is a relevant context. Consider the examples in (19).

(19) a. John began a new book.
b. Jane finished her beer.
c. Bill enjoyed the film.

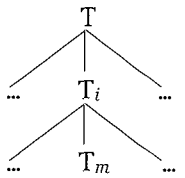
What we can observe here, as Pustejovsky (1995) points out, is that nouns denoting artifacts are 'coerced' into a process or event involving the purpose of that artifact -- reading books, drinking beer, and watching films. But these examples have defeasible interpretations: in suitable discourse contexts, the readings can be replaced by those involving more specific processes (cf. Copestake and Briscoe 1995.) In the examples (20), the original purposes of the nouns *book* and *film* are

turned into unusual purposes relevant to the given context:

(20) a. Fido obviously enjoyed my new book. (reading → eating)
b. Spilberg particularly enjoyed that film. (watching → making)

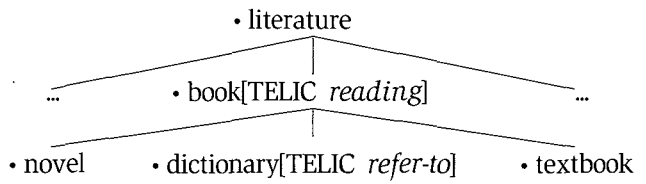
One of the most effective ways of representing these properties of meaning is to accept that all of the concepts in a given context form a conceptual hierarchy or lattice. The overall conception of the conceptual structure of the lexicon can be schematized as in (21):

(21)



In such a conceptual hierarchy, the type T_m, in addition to its own properties, inherits the properties of its supertypes T_i and T as well. In terms of conceptual hierararchies, a subconcept thus inherits all the attributes or concepts of its superconcepts. One thing to note here is that inheritance in the hierarchy is default. For example, the telic role of 'literature' will be *read*, and this conceptual role will be inherited to its subtypes such as *book, novel,* and the like, as represented in (22).

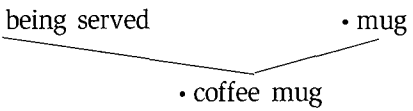
(22)



However, this inheritance rule cannot always be applicable, as hinted in (20). Some of the features can be defeasible in a certain context. *dictionary* also inherits its qualia structure from *book* but has a telic role of *refer-to* rather than *read.* In other words, the default of *read* is overridden. To accommodate this defeasible inheritance system, following Copestake and Briscoe (1995), we adopt the system of default inheritance in which certain conceptal features (particular to one's own) are

defeasible, that is, can be overridden.

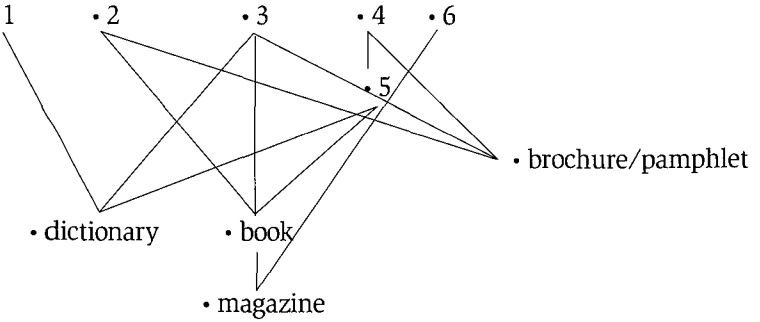
In addition to this issue, there is an issue of 'inequable contribution' problem. As noted, a subconcept, α inherits properties from various superconcepts. But an issue is that not all the inherited properties play an equal role in forming the complex concept α. For example, the concepts 'coffee-serving' and 'mug' make a different contribution to the formation of the concept 'coffee-mug'.

(23) 

To determine what entities denoted by coffee-mug are, we rely more on the concept 'mug'. To solve this inequable contribution problem, we assume the definition of 'base' concept as given in (24):

(24) Base Concept: Concept C is a Base Concept to a complex concept C′ iff C is an immediate superconcept to C′ and play an essential role in determining the entities of C′.

For example, the concepts 'mug' and *ppang* 'bread' are the base concept to the concepts *coffee mug* and *pwunge ppang* 'fish-shaped bread'. In contrast, all of the immediate super-concepts to the complex concepts given in (25) are Base Concepts because they play an equally essential role in determining their entities.

(25) 

(where 1: to be referred to; 2: to be read; 3: written information; 4: paper 5: paper bound with glue; 6: organization)

As seen in (25), the concepts which usually are categorized as concept 'book' are composed of different concepts or properties. In other words, it is customary to take a dictionary, a brochure, or booklet as a kind of book; they are all roughly acceptable as a book as can be seen from their dictionary definitions in (26).

(26) a. A dictionary is a book in which the words and phrases of a language are listed alphabetically...
b. A magazine is a publication with,... (Collins Cobuild English Dictionary 1995)
c. brochure → book → publication, ... (WordNet 1.7 Browser)

To capture, this we assume the definition of 'sister' concept as in (27)[4]

(27) Sister Concept: Two concepts A and B are sister concepts iff they share almost all their immediate superconcepts.

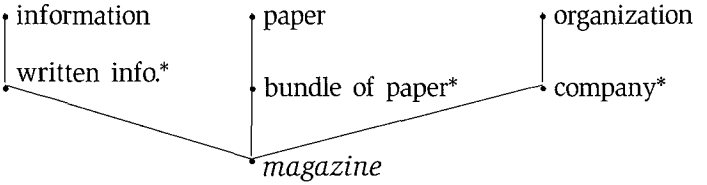
One more notion that is relevant to our discussion is immediate super concept (ISC). As observed so far, ontologically speaking, concepts are thus hierarchically structured. The immediate superconcepts of a concept are the ones that are immediately higher concepts to the concept in such a hierarchy (cf. Priss (1998) and Ganter and Wille (1996)). The hierarchical structures (28)a and (28)b provide examples:

4. As a reviewer pointed, we fully recognize that the definition of 'almost all' needs to be spelled out more clearly. To do so, we can define the notion of Sister Concept in a more rigorous manner (without wording it vaguely with 'almost all') as follows:

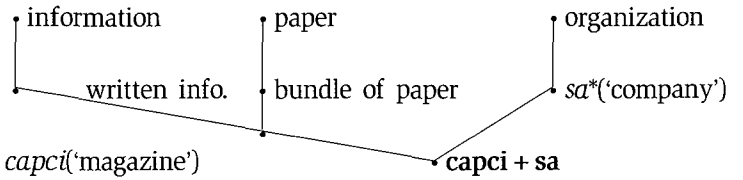
(i) Concepts A and B are a Sister Concept to each other iff for P and Q, such that P is the set of Immediate Super concepts to concept A, and Q to concept B, $P \cap Q \neq \phi$ and $[\#(P \cap Q)/\#(P) \vee \#(P \cap Q)/\#(Q) \geq n - x/n]$, if #(P) or #(Q) is n, and x depends.

In prose, the above definition means that A and B are sister concepts if and only if the sets of their Immediate Super concepts, P and Q share concepts, and furthermore except for the x number of concepts, P and Q share all of their element concepts. Although x will be determined by pragmatics, it might be one, two or maximally three. That is, except one, two or three concepts a pair of sister concepts would be composed of the identical immediate super concepts. See Lee (2000b) for detailed discussion on the definition.

(28) a. Conceptual Structure for \emph{magazine} in English



b. Conceptual Structure for \emph{capci+sa} in Korean



In the hierarchical structure (28)a, the immediate super concept of 'magazine' in English consists of properties (or concepts), 'written information,' 'bundle of paper,' and 'company'. But as seen in (28)b, its Korean counterpart concept 'capci' lacks the property of 'company' or 'organization'. However, the concept 'capci+sa' in Korean has the concepts 'magazine' and 'company' as its immediate superconcepts. Of the two ISCs, the concept 'company' is the only BASE element. This means, in turn, that the sense of 'company' is only available for 'capci+sa' in Korean. The above structures show that *magazine* in English and its Korean counterpart are different conceptually as well as in the way of lexicalizing concepts. This is an integral part in predicting different behaviors of polysemous lexical items in English and Korean.

In sum, we have tried to extend the notion of concepts in an informal way so that we can avoid two conventional problems in representing conceptual structures: inheritance and inequable contribution problem. In particular, in natural language use, entities categorized by a 'sister' concept to a typical concept are also allowed to be named after the name of the typical concept. For example, *dictionary, brochure, pamphlet* and *booklet* are all allowed to be named by the typical concept 'book'.

3. Explaining the Differences within a System of Generative Lexicon

The approach to lexical representation we adopt involves an integration of Pustejovsky (1995) into the typed feature structures of HPSG. Pustejovsky's program basically enriches the semantic representations of words by giving the lexicon greater internal structure, using the mechanisms of *argument-structure, event structure* and *qualia structure*. The structure (29), a reinterpretation of his system within the feature structure of HPSG, is the lexical entry for *bake* in a sentence like *John baked a cake*.

(29)
$$\begin{bmatrix} \langle \text{baked} \rangle \\ \text{CATEGORY verb} \\ \text{ARGSTR} \begin{bmatrix} \text{ARG1 } \boxed{1}\text{NP} \\ \text{ARG2 } \boxed{2}\text{NP} \end{bmatrix} \\ \text{EVENTSTR} \begin{bmatrix} \text{E1 process} \\ \text{E2 state} \\ \text{RESTR } < \\ \text{HEADED E1} \end{bmatrix} \\ \text{QUALIA} \begin{bmatrix} \text{FORMAL bake-result(e2, } \boxed{2}\text{)} \\ \text{AGENTIVE bake-act(e1, } \boxed{1}\text{, } \boxed{2}\text{)} \end{bmatrix} \end{bmatrix}$$

As seen from the structure, ARGSTR contains the conceptual information of each argument. The information encoded in ARGSTR is very similar to the selectional restriction of Chomsky (1965, 1981). EVENTSTR tells us that the accomplishment verb *bake* represents an event comprising a process and a state in which the initial event E1 is focused and thus headed.[5] QUALIA structure is the structure that allows each word to potenitally express multiple (homonymous) meanings, depending on its mode of combination. The QUALIA structure has up to four different types of relation: Constitutive, Formal, Telic, and Agentive (cf. Pustejovsky (1995)).

5. In Pustejovsky's system, verbal events comprise processes, states, and transitions and these events have subeventual structure E1 and E2. The ordering relations of these two events include partial order or overlap.

For example, the noun such as *novel* has the constitution of a 'narrative' and is formally a book. Its telic role is 'reading' and agentive role is 'writing'. This QUALIA structure interacts with ARGSTR and EVENTST. For example, when the noun *novel* combines with the verb *begin* as in *begin a novel*, we could derive either the telic meaning of *reading* or the agentive reading of *writing*.

Based the system of Pustejovsky's (1995) and that of Lee (2000a,b), we propose to enrich semantic information of a lexical entry as in (30) to deal with the issues we have discussed in the previous section. The feature structure in (30), adopting that of HPSG, represents the lexical information of *began* in a sentence like *John began a book*:

(30)
$$\left[\begin{array}{l} \langle \text{began} \rangle \\ \text{CATEGORY}\left[\begin{array}{l}\text{HEAD verb}\\ \text{ARG-STR } \langle \boxed{1}\text{NP}_{\boxed{i}}, \boxed{2}\text{NP}_{\boxed{j}} \rangle\end{array}\right] \\ \text{CONTENT}\left[\begin{array}{l}\text{REST}\left\langle\left[\begin{array}{l}\text{RELATION begin}\\ \text{ACTOR } \boxed{i}\left[\text{FORMAL physobj}\right]\\ \text{UNDERGOER } \boxed{j}\left[\text{FORMAL event} \vee \text{phy-obj}\right]\end{array}\right]\right\rangle \\ \text{EVENTSTR}\left[\begin{array}{l}\text{E1 transition}\\ \text{E2 transition}\\ \text{RESTR overlap}\end{array}\right] \\ \text{ISC}\left[\begin{array}{l}\text{MBRS } \langle \textit{event}, \ldots \rangle\\ \text{BASE } \{ \ldots \}\end{array}\right]\end{array}\right] \\ \text{CONTEXT}\left[\text{BACKGROUND} \mid \text{RESTR}\left\langle\left[\begin{array}{l}\text{RELATION } \psi\\ \text{ARG1 } \boxed{i}\\ \text{ARG2 } \boxed{j}\end{array}\right], \ldots\right\rangle\right] \end{array}\right]$$

The description of ARG-STR has the information about the argument structure of a predicate and the thematic role of its arguments in the sense of Grimshaw (1990). The REST value of the CONTENT contains the semantic specifications of arguments (similar to denotational properties) which we crucially rely on when we try to decide what entity the lexical item denotes. What we can observe here is that the complement of *begin* is marked as an *event* or *physical-object*. And ISC contains the set of sets of properties for the 'immediately supper' concepts. BASE is a set of ISCs which are necessary to determine the

semantic sort of entities by the semantic nucleus. For example, a complex concept like 'coffee mug' consists of its ISCs 'coffee' and 'mug'. In this case, the BASE is the singleton set MUG since a coffee mug is a kind of mug. Furthermore, if ISCs are eventive, then the BASE ISC corresponds to the `Head' subevent of the eventstr.[6]

This system allows us to capture the differences in aspectual verbs like *begin* and its Korean counterpart. Within the system, the structure (31) thus represents the lexical information for *sicakhata* 'begin':

(31)
$$\left[\begin{array}{ll} \text{<시작하다>} & \\ \text{CATEGORY} & \left[\begin{array}{l}\text{HEAD verb}\\ \text{ARG-STR } \langle \boxed{1}\text{NP}_{\boxed{i}}, \boxed{2}\text{NP}_{\boxed{j}} \rangle\end{array}\right] \\ \text{CONTENT} & \left[\begin{array}{ll} \text{REST} & \left\langle \left[\begin{array}{ll}\text{RELATION begin} & \\ \text{ACTOR}_{\boxed{i}} & \left[\begin{array}{l}\text{TYPE animate}\\ \text{FORMAL physobj}\end{array}\right] \\ \text{UNDERGOER}_{\boxed{j}} & \left[\text{FORMAL event - type}\right]\end{array}\right] \right\rangle \\ \text{EVENTSTR} & \left[\begin{array}{l}\text{E1 transition}\\ \text{E2 transition}\\ \text{RESTR overlap}\end{array}\right] \\ \text{ISC} & \left[\begin{array}{ll}\text{MBRS} & \langle \textit{event},\ldots\rangle \\ \text{BASE} & \{\ldots\}\end{array}\right] \end{array}\right] \\ \text{CONTEXT} & \left[\begin{array}{ll}\text{BACKGROUND} \mid \text{RESTR} & \left\langle \left[\begin{array}{ll}\text{RELATION } \psi & \\ \text{ARG1}_{\boxed{i}} & \\ \text{ARG2}_{\boxed{j}} & \end{array}\right], \ldots \right\rangle\end{array}\right] \end{array}\right]$$

The two lexical entries given in (31) show the difference between *begin* in English and its Korean counterpart. They are different in the following sense: the latter says that the word *sicakhata* in Korean can only take an event-denoting expression as its complement; nevertheless the former can take either an event or individual-denoting expression. This difference results in their polysemous behavior observed in section 2. Furthermore, the unbound predicate ψ in the contextual restriction in Korean is also determined by the relevant contextual information. This means that the current way of representing lexical entries returns

6. We accept the viewthat all verbs are associated with at least one event variable.

pragmatic meaning to the field of pragmatics unlike Pustejovsky (1995). In short, the present way is more flexible or competent in accounting for the polysemous behaviors of the word *begin*.[7]

Then, let us see how this system generates the lexical entry for English *magazine* and its Korean counterpart *capci* as given in (32) a and (32)b.

(32) a.
$$\left[\begin{array}{l}\text{<magazine>}\\ \text{CONT}\left[\begin{array}{l}\text{RESTR}\left\langle\left[\begin{array}{l}\text{REL book}\\ \text{INST } i\end{array}\right],\left[\begin{array}{l}\text{REL bundle-of paper}\\ \text{INST } i\end{array}\right],\left[\begin{array}{l}\text{REL written-inf}\\ \text{INST } i\end{array}\right],\left[\begin{array}{l}\text{REL regularly-published}\\ \text{INST } i\end{array}\right]\right\rangle\\ \left[\begin{array}{l}\text{REL, organization}\\ \text{INST } i\end{array}\right]\\ \text{ISC}\left[\begin{array}{l}\text{MBRS } \langle\text{written info, company, bundle of paper,...}\rangle\\ \text{BASE } \{\text{written info, bundle of paper,...}\}\end{array}\right]\end{array}\right]\end{array}\right]$$

b.
$$\left[\begin{array}{l}\text{<잡지>}\\ \text{CONT}\left[\begin{array}{l}\text{RESTR}\left\langle\left[\begin{array}{l}\text{REL book}\\ \text{INST } i\end{array}\right],\left[\begin{array}{l}\text{REL bundle-of paper}\\ \text{INST } i\end{array}\right],\left[\begin{array}{l}\text{REL written-inf}\\ \text{INST } i\end{array}\right],\left[\begin{array}{l}\text{REL regularly-published}\\ \text{INST } i\end{array}\right]\right\rangle\\ \text{ISC}\left[\begin{array}{l}\text{MBRS } \langle\text{written info, bundle of paper,...}\rangle\\ \text{BASE } \{\text{written info, bundle of paper,...}\}\end{array}\right]\end{array}\right]\end{array}\right]$$

7. In the present system, it is plausible to account for the cancellability difference between English and Korean accomplish-verbs in the past tense in terms of lexical difference because it is implausible to treat the past tense for either language differently (see section 2.4). The event denoted by *kkay-ta* 'break' in Korean consists of two sub-events of breaking process and resultant state of being broken. This information is encoded both in EVENTSTR and in ISC. In this sense, *break* in English is the same as its Korean counterpart *kkay-ta*. However, they differ from each other in the following sense: *break* in English denotes an event which has the resultant state as the head sub-event. On the contrary, the event denoted by *kkay-ta* in Korean is underspecified in terms of such headness. In other words, the EVENTSTR of *kkay-ta* in Korean doesn't contain the information about headness or is underspecified in terms of headness; that of *break* in English bears the resultant state as the 'head' subevent. The presence or absence of the explicitly specified information about a head subevent in the EVENTSTR determines the accomplishment-cancellability. Verbs like break in the past tense always deliver the non-cancellability reading because the past tense modifies the head subevent (resultant state).

The only difference between the lexical entries for English *magazine* and its Korean counterpart *capci* 'magazine' is that the concept of 'magazine' contains the concept 'organization' while the Korean counterpart doesn't. This difference is responsible for the polysemous incongruity observed in section 2. As shown in section 2, a bound morpheme *-sa* 'company' is required to attach to the end of *capci* so as to carry the meaning of a magazine company. In traditional Korean dictionaries, *capci+sa* is registered as an independent word. Hence, *capci+sa* in Korean is conceptually similar to the concept of English 'magazine' in (32)a. The lexical entry for the Korean word *capci+sa* can be given as in (33):

(33)
$$\left[\begin{array}{ll} \langle \text{잡지사} \rangle & \\ \text{CONT} & \left[\begin{array}{ll} \text{RESTER} & \left\langle \left[\begin{array}{l}\text{REL organization}\\ \text{INST i}\end{array}\right], \left[\begin{array}{l}\text{REL publish}\\ \text{ARG1 i}\\ \text{ARG2 j}\end{array}\right], \left[\begin{array}{l}\text{REL regular-magazine}\\ \text{INST j}\end{array}\right] \right\rangle \\ \text{ISC} & \left[\begin{array}{l}\text{MBRS } \langle \text{magzine, compang,...} \rangle\\ \text{BASE } \{ \text{company,...} \}\end{array}\right] \end{array}\right] \end{array}\right]$$

The main difference between the lexical entries given in (32)a and (33) is that in the former, three concepts (e.g. 'organization,' 'information,' 'physical object') are all equally available while in the latter only the concept 'organization' is available because it is represented as the only BASE ISC. This difference prevents *capci+sa* in Korean from being interpreted as the content of the magazine or the magazine itself, the physical object as the following data show:

(34) a. *John-nun capcisa-lul ilk-ess-ta.
John-Top magazine co.-Acc read-pst-Dec
Intended Interpretation: 'John read the magazine.'

b. *John-nun capcisa-lul kkalkoanc-ass-ta.
John-Top magazine co.-Acc sit on-pst-Dec
Intended Interpretation: 'John sat on the magazine.'

c. John-nun capcisa-lul kopalhay-ess-ta
John-Top magazine-co.-Acc accuse-pst-Dec
Intended Interpretation: 'John accused the magazine.'

In sum, the concept of 'magazine' in English and that of *capci* in Korean are not completely identical. However the concept of *capci+sa* is much different from that of 'magazine' in English in that the former categorizes the set of magazine companies while the latter covers three categories, namely magazines as physical entities, information contained in the magazines, and magazine companies simultaneously. Such incongruities in concept and lexicalization between two languages result in the different polysemous behaviors. That is, rarely does *capci* in Korean carry the sense of a magazine company. This task is taken care of by the independent lexical item *capci+sa* in Korean.

4. Concluding Remarks

So far, we have provided an account of lexical and conceptual differences between English and Korean by extending the formal concept theory of Ganter & Wille's (1996), and Priss (1998). With the newly defined notions of 'sister concept' and 'base concept', we proposed an alternative way of representing lexical entries to Pustejovsky's (1995) analysis. The extended formal concept theory can overcome the two traditional problems, namely, 'Inheritance problem' and 'Inequable Contribution' problem. Furthermore, compared with Pustejovsky (1995), the current way of representing lexical entries is free from most of Pustejovsky's problems, and is more efficient in reflecting lexical differences observed above between English and Korean. However, the current analysis also has its own inherent problems: in terms of the manner of representing lexical entries, it is not mentioned how Immediate Super Concepts (to a concept) are determined. With regard to complex concepts and their component superconcepts, We have provided no discussion about what it means for a complex concept to consist of some concepts. It is obvious that the set intersection is not only a possible link applicable to superconcepts (to a complex concept). In the case of complex concept

'book', the set intersection operation over its superconcepts 'written information' and 'bound paper with glue' cannot bring the complex concept into being. Instead, the relation 'hold' can. All these problems and other relevant ones have to await further study.

References

Bartsch, Renate. (1998) *Dynamic Conceptual Semantics*, Stanford, CSLI Publication.

Chomsky, Noam. (1965) *Aspects of the Theory of Syntax*, MA, MIT Press.

Chomsky, Noam. (1981) *Lectures on Government and Binding*, Dordrecht, Kluwer Academic Publishers.

Copestake, Ann and Ted Briscoe, (1995) 'Semi-productive Polysemy and Sense Extension,' *Journal of Semantics*, 12, 15-67.

Ganter, Bernhard. and Rudolf. Wille (1996) *Formal Concept Analysis: Mathematical Foundation*, Berlin, Springer.

Grimshaw, Jane. (1990) *Argument Structure*, MA, MIT Press.

Hofmann. T. R. (1993) *Realms of Meaning: An Introduction to Semantics*, Longman Publishing.

Lakoff, George. (1988) *Cognitive Semantics, in Meaning and Mental Representation.* Eckman, F. R. et al., 119-154, London, Plenum Press.

Langacker, Ronald. (1998) Cognitive Grammar: A Basic Introduction, Handout read at Cognitive Linguistics Conference at KNU.

Lascarides, Alex. and Ann Copestake, (1999) 'Default representation in constraint-based frameworks,' *Computational Linguistics*, 25:1, 55-105.

Lee, Yae-Shiek. (2000a) 'Lexical Mismatches between English and Korean: with Particular Reference to Polysemous Nouns and Verbs,' *The Language and Information.* 4.1.

Lee, Yae-Shiek. (2000b) Incongruities in the Lexicons of English and Korean, Paper read at ICKL-2000, Prague.

Levin, Beth (1993) *English Verb Classes and Alternations: A Preliminary Investigation*, The University of Chicago Press.

Priss, Uta E. (1998) The Formalization of WordNet by Methods of Relational Concept Analysis, in WordNet: an electronic lexical database, 179-196. MA, MIT Press

Pustejovsky, James. (1995) *The Genetative Lexicon*, MA, the MIT Press
Talmy, Leonard. (1985) 'Lexicalization patterns: semantic structure in lexical forms,' in *Language Typology and Syntactic Description*, vol. 3, ed. by Spophen, T., 57-149, Cambridge, England, Cambridge University Press.
Wienold, Götz. (1995) 'Lexical and Conceptual Structures in Expressions for Movement and Space: with reference to Japanese, Korean, Thai, and Indonesian as compared to English and German,' in *Lexical Knowledge in the Organization of Language* eds., by Egli, U. et al., 301-340, Amsterdam, John Benjamins.

ABSTRACT

Capturing Some Lexical Differences between English and Korean: A Generative Lexicon Approach

Yae-Sheik Lee · Jong-Bok Kim

Cross-linguistic lexical differences pose great obstacles to natural language processing applications such as multilingual information retrieval and machine translation. This paper shows that most of such differences can be attributed to differences in the information of lexical entries. It begins by reviewing lexical mismatches between English and Korean. The paper then proposes a conceptual system that adopts a multiple inheritance system. Second, to deal with the meaning of an expression, we adopt a cognitive semanticists' view (i.e., Langaker 1998). This view assumes that meaning resides in concept. According to formal concept theories like Ganter & Wille (1996), and Priss (1998), a concept can be further specified as a set of attributes or properties and a set of objects. We extend such a formal concept theory within the constraint-

based framework of HPSG (Head-driven Phrase Structure Grammar). We also show a formalism of representing the information of lexical entries incorporating Pustejovsky's (1995) generative lexicon into the framework of Head-driven Phrase Structure Grammar that allows tight interactions among various grammatical components such as syntax, semantics, and pragmatics.

Yae-Sheik Lee
Deparment of English Education
Kyungpook National University
Sangyeok-dong 1370, Buk-gu
Daegu 702-701, Korea
E-mail : yaesheik@kyungpook.ac.kr

Jong-Bok Kim
School of English
Kyung Hee University
1 Hoegi-dong, Dongdaemun-gu
Seoul 130-171, Korea
E-mail : jongbok@khu.ac.kr